Lecture Outline

Thus, the lecture will contain:

1. Introduction
2. How to frame the business problem
3. How to transfer it to a problem which can be solved with analytics methods
4. Data identification and prioritisation, data collection and data harmonisation
5. **Identification of problem solving approaches and appropriate tools (not only R even though this is important)**
6. How to set up and validate models
7. The deployment of a model
8. Model lifecycle
9. Some words about soft skills needed by statistical and mathematical professionals

Chapter 5
Identification of problem solving approaches and appropriate tools

5.1 Identify available problem solving approaches / methods

Where to start with identifying the appropriate method?

5.1 Identify available problem solving approaches / methods

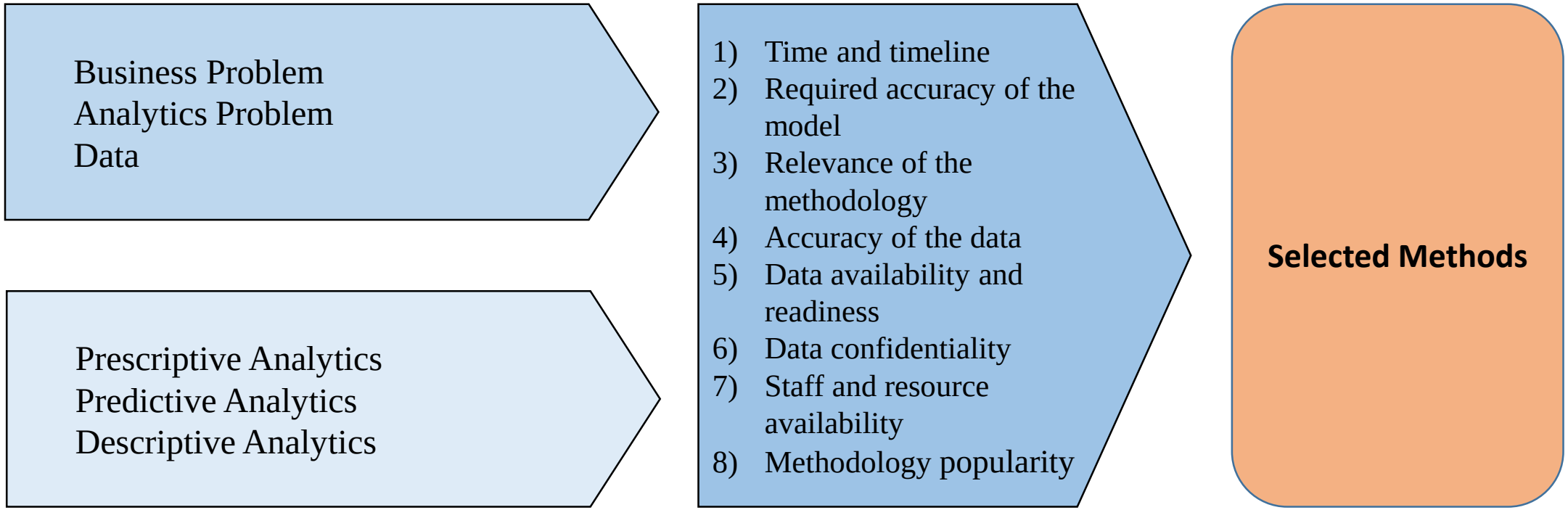
Where to start with identifying the appropriate method:

- Definition of the business problem: What we want to answer
- Definition of the analytics problem: How we want to solve it analytically
- Data: Which data are available and in which quality and granularity

5.1 Identify available problem solving approaches / methods

Classification	Questions	Examples from Business	Purpose
Prescriptive	What is the best outcome? What if?	Optimisations Scenario testing	• Evaluates and determines new ways to operate • Targets business objectives • Balances all constraints
Predictive	What could happen? What is happening next? Why is this happening?	Statistical modelling Forecasting	• Predicts future probabilities and trends • Finds relationships in data not readily apparent with traditional analysis
Descriptive	What happened? How many, how often? What action is needed?	Standard and ad hoc reports Queries Alerts	• Prepares and analyzes historical data • Identifies patterns from samples for reporting of trends

5.1 Identify available problem solving approaches / methods



5.1 Identify available problem solving approaches / methods

1) Time and timeline

- Typically, tight deadlines and the work should be completed “yesterday”
- Methods are chosen with respect to the ability to perform the analysis within the timeframe i.e. this means often no experiments and no trials for new methods
- Methods which can be run very quickly and with which one is familiar are typically chosen

5.1 Identify available problem solving approaches / methods

Example

In Human Resource (HR) analytics where one wanted to determine the key success factors of HR management to the business and the people working in business only the influence of single factors have been analysed without taking into consideration interdependencies of such key success factors e.g. the level of education and the attended trainings.

5.1 Identify available problem solving approaches / methods

2) Required accuracy of the model

- Level of aggregation / granularity of a model influence the accuracy of the results
- Additionally, it depends on the quality and readiness of data
- If the quality and accuracy of data is not given, very accurate and granular models give no additional value

5.1 Identify available problem solving approaches / methods

Example

Old age homes analytics in a certain canton:

- Analytics per financial line item has not given any comparable results as the line items have been too inhomogeneous and the cut offs differs from one old age home to another
- But applying methods on aggregated data e.g. maintenance costs for buildings have produced useful results

5.1 Identify available problem solving approaches / methods

3) Relevance of the methodology

- The result of the business problem statement and the analytics problem framing often point the direction to a methodology
- It is important to understand the goal of the analysis and if this is descriptive, predictive or prescriptive

5.1 Identify available problem solving approaches / methods

Example

In the analysis of the customer of the luxury store the results out of the business problem statement was that they do not have any customer insights at all.

Further, the stage of management in data analytics awareness and experience was low developed.

Thus, one has to start with descriptive analytics to give them first the fundamentals of insights and to educate them in the basic understanding of analytics results.

5.1 Identify available problem solving approaches / methods

4) Accuracy of the data

- Accuracy of data influences and restrict the pool of possible methods
- Analysis of the accuracy on a granular level versus the accuracy on an aggregated level of the data

5.1 Identify available problem solving approaches / methods

Example

Hospital optimization in two different hospitals:

- In one hospital granular data for each and every time point of action e.g. begin anesthesia, entering the operating room, begin of surgery and so on was available and because electronical real-time recording the data have been of good quality.
- In another hospital the data have been recorded manually and sometimes with huge delays and thus, the data have been very imprecise.

Thus, in the first case we could granular time series models and taking into consideration seasonalities whereas in the second case we had to work more with regression analysis and smoothing out inconsistencies.

5.1 Identify available problem solving approaches / methods

5) Data availability and readiness

- Depending on the available data the methods have to be chosen
- Additionally, it can happen that data are available
  - but not easy accessible or
  - only available in very poor quality or
  - available but no information about the data are available and thus, the quality and relevance of the data is not known at all

5.1 Identify available problem solving approaches / methods

Example

Pricing of products for an accident insurer:

In the mandatory accident insurance area (UVG) many data are typically available and back more than 25 years. Thus, various different methods from descriptive methods, trend estimations to generalized linear models can be used.

For special products or new launched products e.g. supplemental accident insurance for retired people only few or no data are available.

5.1 Identify available problem solving approaches / methods

Example

The determination of the credit risk of the customer of a telecom company:

Data are available but stored in a datawarehouse on old IT infrastructure such that the extracting the data out of the system needs first a lot of IT programming and setting accessibility to this database and second, the port for extracting the data has limited access capacity such that it would take several weeks for only extracting the data

5.1 Identify available problem solving approaches / methods

6) Data confidentiality

- Depending how confidential data is i.e. internally or by data protection law, data is only available highly anonymized and/or aggregated

5.1 Identify available problem solving approaches / methods

Example

Patient data from which one could reconstruct the detailed disease or treatment reason have to be anonymized in a way that for the analytics professional it is not possible to identify the person. I.e. no names, no date of birth, sometimes even no gender, no address, only certain fields which contains indications about the disease are available.

5.1 Identify available problem solving approaches / methods

7) Staff and resource availability

- For different areas there are different experts:
  - Experts in statistics
  - Experts in Operations Research
  - Experts in pricing
  - and so on…

And data scientists / data analytics professional at all are scarce in the market.

- There may be also restrictions in resources due to available licenses for an analytics software especially, if one have to work on premise of your customer

5.1 Identify available problem solving approaches / methods

Example
Hospital optimization: Requirement of a person who is familiar with System Dynamics and Discrete Event modelling and a person who is familiar with time series modelling and simulations:

Step 1: descriptive analytics thus, analysis of the hospital data

Step 2: building up the model
Step 3: for optimization purpose there was the need of simulated patient
data which have given the time developments and included long-time trends
Finally, the project had a delay of several weeks just because the corresponding experts had not been available on the agreed project start and some short-cuts had to be done such that one could finish in time.

5.1 Identify available problem solving approaches / methods

8) Methodology popularity

- Some methods are well known in the market and “everybody is applying them”
- Sometimes the customer is demanding a certain analysis / method to apply
- It is in the responsibility of the data analytics professional to apply the most accurate method and not only the most popular / demanded one

5.1 Identify available problem solving approaches / methods

Example

A Human Resource (HR) department requested a “structural equation modeling”*) analysis because they have read articles about this method. Already in the first meetings one could see that this was not the appropriate method for solving their problem i.e. measuring the HR management influence to the success of internal projects. Thus, we applied the “classical” approaches like regression but the most important point was the selection of the variables (and not the model).

*) this is a method to test a conceptual model where the observation error can be isolated especially in setting where a direct measurement is not possible.

It is case study time